雑感　「データの分析」とΣ計算

■　数学Ⅰの「データの分析」では，平均値，分散，共分散，相関係数などの値が定義され，その値が分析に活用される．

しかし，総和Σの記号が数学Bで扱われるため，Σを用いない長ったらしい式でこれらの値が表記される．その長い表記は初学者には分かり易いという側面を持つ一方，式の書き出しは煩雑で，Σの持つ性質を簡単に使い記述して行く上では，極めて不便である．

■　数学Bまでを試験科目に含めている大学では，Σを用いた表記での出題が普通にある．

例えば2016年の一橋大，広島大の問題ではそれぞれ

「（略）$2n$ 個の実数 $a_1,\ a_2,\ \cdots,\ a_n,\ b_1,\ b_2,\ \cdots,\ b_n$ について，

$a=\dfrac{1}{n}\displaystyle\sum_{k=1}^{n}a_k,\ b=\dfrac{1}{n}\sum_{k=1}^{n}b_k$ とすると，$\displaystyle\sum_{k=1}^{n}(a_k-a)(b_k-b)=\sum_{k=1}^{n}a_kb_k-nab$

が成り立つことを示せ．（略）」，

「（略）$f(a)=\dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(x_k-a)^2$ とする．（略）」とある．

この表記だと，$\displaystyle\sum_{k=1}^{n}(a_k-a)(b_k-b)=\sum_{k=1}^{n}a_kb_k-a\sum_{k=1}^{n}b_k-b\sum_{k=1}^{n}a_k+nab$，

$f(a)=\dfrac{1}{n}\displaystyle\sum_{k=1}^{n}{x_k}^2-\dfrac{2a}{n}\sum_{k=1}^{n}x_k+\dfrac{n}{n}a^2$ といった式変形がすぐに従う．

■　センター試験だとこうはいかない．2017年の追試験．

〔3〕　A組 m 人とB組 n 人の生徒に対して行ったテストの得点を

A組　$x_1,\ x_2,\ \cdots,\ x_m$

B組　$y_1,\ y_2,\ \cdots,\ y_n$

と書く。各組の平均点を $\overline{x}$，$\overline{y}$，分散を ${S_A}^2$，${S_B}^2$ とする。また，A組とB組を合わせた $(m+n)$ 人の得点の平均点を $\overline{w}$，分散を S^2 とする。これらの間に一般に成り立つ関係について調べる。

A組の得点と $\overline{w}$ の差の2乗の和

$$(x_1-\overline{w})^2+(x_2-\overline{w})^2+\cdots+(x_m-\overline{w})^2$$

を，$\overline{x}$，${S_A}^2$，$\overline{w}$ を用いて表すと $\boxed{\text{チ}}$ である。ただし，${S_A}^2$ は

$${S_A}^2=\frac{1}{m}({x_1}^2+{x_2}^2+\cdots+{x_m}^2)-(\overline{x})^2$$

を，$\overline{x}$，${S_A}^2$，$\overline{w}$ を用いて表すと $\boxed{\text{チ}}$ である。ただし，${S_A}^2$ は

で計算できる。$\boxed{\text{チ}}$ に当てはまるものを，次の⓪～③のうちから一つ選べ。

⓪　${S_A}^2+(\overline{x})^2+(\overline{w})^2$　　①　${S_A}^2+(\overline{x}-\overline{w})^2$

②　$m{S_A}^2+m\{(\overline{x})^2+(\overline{w})^2\}$　　③　$m{S_A}^2+m(\overline{x}-\overline{w})^2$

A組とB組の生徒を合わせた $(m+n)$ 人の得点の分散 S^2 は $\boxed{\text{ツ}}$ に等しい。$\boxed{\text{ツ}}$ に当てはまるものを，次の⓪～④のうちから一つ選べ。

⓪　$\dfrac{m{S_A}^2+n{S_B}^2+(m+n)\{(\overline{x}+\overline{y})^2-(\overline{w})^2\}}{m+n}$

①　$\dfrac{m{S_A}^2+n{S_B}^2-(m+n)\{(\overline{x})^2+(\overline{y})^2-(\overline{w})^2\}}{m+n}$

②　$\dfrac{m{S_A}^2+n{S_B}^2-\{m(\overline{x})^2+n(\overline{y})^2\}+(m+n)(\overline{w})^2}{m+n}$

③　$\dfrac{m{S_A}^2+n{S_B}^2+m(\overline{x})^2+n(\overline{y})^2+(m+n)(\overline{w})^2}{m+n}$

④　$\dfrac{m{S_A}^2+n{S_B}^2+m(\overline{x})^2+n(\overline{y})^2-(m+n)(\overline{w})^2}{m+n}$

長々とした式が登場するが，A組の得点と $\overline{w}$ の差の2乗和は

$\displaystyle\sum_{k=1}^{m}(x_k-\overline{w})^2=\sum_{k=1}^{m}{x_k}^2-2\overline{w}\sum_{k=1}^{m}x_k+m(\overline{w})^2$ ……①　であり，

ただし書きの式から $\displaystyle\sum_{k=1}^{m}{x_k}^2=m{S_A}^2+m(\overline{x})^2$ であるから，①は

$m{S_A}^2+m(\overline{x})^2-2\overline{w}\cdot m\overline{x}+m(\overline{w})^2=m{S_A}^2+m(\overline{x}-\overline{w})^2$ となり，③となる．

ここで導いた式は後半の設問の決定的なヒントであり，

$(m+n)S^2=\displaystyle\sum_{k=1}^{m}(x_k-\overline{w})^2+\sum_{k=1}^{n}(y_k-\overline{w})^2=m{S_A}^2+m(\overline{x}-\overline{w})^2+n{S_B}^2+n(\overline{y}-\overline{w})^2$

$=m{S_A}^2+n{S_B}^2+m(\overline{x})^2+n(\overline{y})^2-2(m\overline{x}+n\overline{y})\overline{w}+(m+n)(\overline{w})^2$

$=m{S_A}^2+n{S_B}^2+m(\overline{x})^2+n(\overline{y})^2-2(m+n)\overline{w}\cdot\overline{w}+(m+n)(\overline{w})^2$

$=m{S_A}^2+n{S_B}^2+m(\overline{x})^2+n(\overline{y})^2-(m+n)(\overline{w})^2$ より，④となる．

言うまでもなく，Σが威力を発揮している．

■　最後の内容は，当雑感128「合併2集団の分散」で扱った内容である．難易度が高い問題中，このヒントは秀逸である．とは言え，表記方法がそこで触れたものとも異なり，妖怪の面目躍如である．

2017年08月01日